DOCUMENT RESUME

ED 361 370 TM 020 460

AUTHOR	Silver, Edward A.; Lane, Suzanne
TITLE	Balancing Considerations of Equity, Content Quality, and Technical Excellence in Designing, Validating and Implementing Performance Assessments in the Context of Mathematics Instructional Reform: The Experience of the QUASAR Project.
SPONS AGENCY	Ford Foundation, New York, N.Y.
PUB DATE	93
CONTRACT	890-0572
NOTE	60p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Cognitive Tests; Cultural Differences; Curriculum Development; *Curriculum Enrichment; Disadvantaged Youth; Economically Disadvantaged; Educational Assessment; Educational Discrimination; Educationally Disadvantaged; Educational Quality; Elementary School Students; *Equal Education; Ethnic Groups; Intermediate Grades; Junior High Schools; Junior High School Students; Mathematics Achievement; *Mathematics Instruction; *Middle Schools; Minority Group Children; *Test Use
IDENTIFIERS	Middle School Students; Open Ended Questions; *Performance Based Evaluation; QUASAR Cognitive Assessment Instrument; *QUASAR Project (Mathematics Education)

ABSTRACT

Issues of educational equity and quality are explored in the context of the Quantitative Understanding: Amplifying Student Achievement and Reasoning (QUASAR) project, a national educational reform project aimed at fostering and studying the development and implementation of enhanced mathematics instructional programs for students attending middle schools in economically disadvantaged communities. Currently operating at six school sites across the country, QUASAR is based on the premise that current poor mathematical achievement by disadvantaged students is not the result of lack of ability, but rather the result of educational practices that have blocked students from meaningful experiences. Assessment in QUASAR is embedded in the larger project. A major component is the QUASAR Cognitive Assessment Instrument (QCAI), which assesses student performance on open-ended tasks involving mathematical problem solving, reasoning, and communication, with a focus on efforts to obtain content appropriateness, technical measurement quality, and equity. Evidence of the validity of the QCAI for diverse linguistic and cultural groups is reviewed. The experience of QUASAR should be valuable to others interested in program and assessment improvement. One table and two figures illustrate the discussion. An appendix presents general rubric components. (Contains 58 references.) (SLD)

ED 361 370

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
K. MALOY
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Balancing Considerations of Equity, Content Quality, and Technical Excellence in Designing, Validating and Implementing Performance Assessments in the Context of Mathematics Instructional Reform: The Experience of the QUASAR Project

Edward A. Silver and Suzanne Lane

Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260

Preparation of this paper was supported by a grant from the Ford Foundation (grant number 890-0572) for the QUASAR project. Any opinions expressed herein are those of the authors and do not necessarily represent the views of the Ford Foundation.

TM020460

BEST COPY AVAILABLE

Balancing Considerations of Equity, Content Quality, and Technical Excellence in Designing, Validating and Implementing Performance Assessments in the Context of Mathematics Instructional Reform: The Experience of the QUASAR Project

Reports by the National Academy of Sciences (National Research Council, 1989) and the National Council of Teachers of Mathematics (1989) have focused the attention of educational practitioners and policy makers on mathematics education reform in the United States. These reports, and others like them, have been issued at a time when concerns have been expressed publicly about evidence that American students are unable to perform at acceptable levels in mathematics and about the implications of this low level of performance for the nation's long-term economic competitiveness in an increasingly technological world. The reports specify new goals -- sometimes referred to as "world class standards" for mathematics education --and provide new descriptions of mathematical proficiency, using terms like reasoning, problem solving, communication, conceptual understanding, and mathematical power. Not only do these reports offer an expanded view of mathematical proficiency but they also indicate that high-level mathematical goals and outcomes should be expected of all students (Silver, 1992b).

There is an important equity dimension to the current interest in upgrading the quality of mathematics education. Data from numerous research studies have revealed the intellectually and academically impoverished nature of the content and instructional style in mathematics classes attended by most American students, but especially by students who are members of racial and ethnic minority groups or those who live in poverty. These data, coupled with demographic predictions that these groups, who are currently the least well served by our educational system, will comprise an increasingly large proportion of American society in the coming decades, have raised concerns about equity as it relates to the mathematics education

reform agenda.

The potential of this country's cultural diversity has not been fully developed, nor have all children been given equal opportunity to learn mathematics and other school subjects that allow access to employment and further education. The centrality of mathematics and the disastrous implications of this situation were both noted by the National Research Council in Everybody Counts, a report to the nation on the state of mathematics education:

> Because mathematics holds the key to leadership in our information-based society, the widening gap between those who are mathematically literate and those who are not coincides, to a frightening degree, with racial and economic categories. We are at risk of becoming a divided nation in which knowledge of mathematics supports a productive, technologically powerful elite while a dependent, semiliterate majority, disproportionately Hispanic and Black, find economic and political power beyond reach. Unless corrected, innumeracy and illiteracy will drive America apart. (1989, p. 14)

Thus, although mathematics education reform appears to focus on a relatively narrow, though important, slice of the educational pie, it can be seen to be related to a more general agenda of equity and social justice. Demographic trends indicate that the continued underinvestment in the education of the poor, disproportionate numbers of whom are members of racial or ethnic minority groups, will exacerbate the current achievement gaps between groups in this society and between U. S. students and their counterparts in other industrialized societies (Pallas, Natriello & McDill, 1989).

Issues related to student assessment have also been prominent in mathematics education reform discussions. Considerations of how to assess students' attainments with respect to a new vision of mathematical proficiency and how to assess improvements that may result from curricular and instructional reforms that might be undertaken are a natural consequence of the

current interest in educational reform. In fact, alternatives to conventional forms of mathematics assessment (e.g., standardized multiple-choice tests) have been prominent in much of the conversation about mathematics education reform (Silver, 1992a). Although there is considerable rhetoric about alternative forms of mathematics assessment and there are some impressive prototypes of new forms of assessment tasks (National Research Council, 1993), there are relatively few examples of large-scale assessments that have been developed to measure new forms of mathematical proficiency, and even fewer that have ample reliability and validity evidence to support the use of the assessment and the interpretation of the derived scores. Thus, current efforts to implement the mathematics reform agenda provides an arena in which a wide variety of technical issues associated with the design, implementation and validation of performance assessments can be examined.

This paper focuses on student assessment within the QUASAR project, which is aimed at increasing equitable student access to high quality mathematics instruction. Because of the nature of the project, its assessment development and implementation efforts have been influenced by the need to balance concerns about mathematical content quality with those about technical measurement, equity and fairness. We begin with a fairly brief consideration of the current situation in mathematics education, after which a general description of QUASAR is provided. In both sections, particular attention is paid to many of the equity-related aspects of the project and to the context in which it operates. Finally, the nature and role of assessment in the project is discussed, and some specific, equity-related details are given regarding the development, implementation and validation of one of its assessment instruments.

The QUASAR Project: Background and Context

QUASAR (Quantitative Understanding: Amplifying Student Achievement and Reasoning)

is a national educational reform project aimed at fostering and studying the development and implementation of enhanced mathematics instructional programs for students attending middle schools in economically disadvantaged communities. Launched in the Fall of 1989, and currently operating at 6 school sites dispersed across the United States, QUASAR aims to demonstrate that students in these communities can and will learn a broader range of mathematical content, acquire a deeper and more meaningful understanding of mathematical ideas, and demonstrate proficiency in mathematical reasoning and solving appropriately complex mathematics problems. An important aspect of QUASAR is its extensive research and evaluation effort that seeks to document the goals and plans of the local efforts, the nature of the implementation of the plans in particular schools and classrooms, and the impact of the implementation on teachers and students.

Before presenting the details of the project and its approach to student assessment, it is useful to consider the general educational context in which the project has been undertaken. This brief examination of context will focus on data regarding mathematical proficiency and conventional instructional and school organizational practices that affect the mathematics achievement of students, with special attention to a few equity issues of special relevance to the QUASAR project. It is beyond the scope of this paper to provide a full presentation of equity issues as they pertain to student performance in mathematics, or even as they relate specifically to the project. A more complete report of general issues is provided by Secada (1992), and a more extensive discussion of these issues as they pertain to QUASAR is given by Silver, Smith and Nelson (in press).

A Crisis of Participation and Performance in Mathematics

There is a widely-recognized crisis in mathematics education related to low rates of student participation, insufficient student access to quality mathematics instruction and inadequate student performance in mathematics. With respect to participation, data available from the recent NAEP mathematics assessments (Dossey, Mullis, Lindquist, & Chambers, 1988; Mullis, Dossey, Owen, & Phillips, 1991) indicate that too few students are electing to take advanced mathematics courses and studying mathematics throughout their high school years. The NAEP data suggest that, for the nation as a whole, only nine of every one hundred graduating high school students completes four years of college preparatory mathematics and is thereby prepared adequately for the study of calculus in college. In disadvantaged urban communities, the participation rate in advanced mathematics courses is even worse: only five of every one hundred students completes four years of college preparatory mathematics. In urban schools serving economically disadvantaged communities, students take very little mathematics at all. In fact, four of five students take no math beyond the minimum required for graduation, which may be as little as two years of pre-algebra coursework. Furthermore, although the college-attending rates of minority and majority students are almost identical for students who have taken algebra and geometry in high school (Pelavin & Kane, 1990), NAEP data indicate that less than half the students in urban schools take any mathematics beyond one year of algebra, and one in five do not study algebra at all.

With respect to performance, results of national and international assessments have provided sobering statistics regarding the impoverished state of American students' mathematical proficiency, especially with respect to complex tasks and problem solving (Bourque & Garrison, 1991; Robitaille & Garden, 1989). Not only are there too few American students performing a the highest levels on these assessments, but there are too few females, ethnic minorities or students from poor communities in the group of high-performing students. In fact,

the vast majority of students are achieving at levels substantially below international standards.

From an equity perspective, not all the news concerning mathematics performance has been gloomy. For example, it has been reported that minority students have narrowed the achievement gap in standardized test performance (Congressional Budget Office, 1987) and on NAEP (Mullis, Owen & Phillips, 1990) over the past two decades, and that minority students have improved at a faster rate than their white counterparts. Secada (1992) carefully reviewed the evidence that has been presented to support these claims, and he concluded that the story is much less clear than it appears at first glance. For example, the validity of conclusions drawn from the achievement test data base is compromised by the "Lake Wobegon" effect reported by Cannell (1988), who found that virtually all states and school districts used the same type of norm-referenced, achievement test scores to report that their students were performing above the national average. The NAEP data are likewise unclear in their implications, since the pattern of change for various minority groups does not appear to be the same over the past few assessment administrations. Matthews, Carpenter, Lindquist and Silver (1984) reported a consistent pattern of gains for minority students on all types of questions (e.g., knowledge, skills, understanding, and problem solving) on the NAEP assessment between 1978 and 1982; whereas, the gains reported for the 1986 assessment were limited to lower-level questions and primarily to African American students and not Hispanic students. Moreover, the magnitude of the observed changes in NAEP has been fairly small.

Despite uncertainty about the uniformity and magnitude of the changes in performance of minority students, the evidence does suggest that some improvements have occurred. The good news is that the observed improvements almost certainly indicate that the additional financial support, made available through Chapter 1 to schools serving economically disadvantaged communities, has been used to advantage and that the students in these programs

have learned what they have been taught (Birman, 1987). Nevertheless, despite the positive outcome of reducing intergroup performance differences, the NAEP gains have generally come from improved performance only on those portions of tests related to factual knowledge and basic calculation skills; little change has been found for portions of the test measuring higher-level mathematical outcomes (Secada, 1992b). The lack of improvement on more complex mathematics tasks suggests that available instruction has been focused primarily on low-level objectives, and this inference is supported by data, which are discussed in the next section of the paper, regarding the low-level emphasis of instruction in the lower tracks of many high schools, especially those in poor communities.

From the perspective of mathematics education, the above data collectively point to the need to improve mathematics course enrollment and mathematics achievement for all American students, with a special emphasis on increasing in poor communities the level of students' participation and performance in a mathematics sequence that takes them at least as far as Algebra and Geometry. Since the trajectory for high school participation and performance in mathematics is set prior to ninth grade (Oakes, 1990b), it is imperative that these issues be addressed in middle school mathematics programs, and this is being done in the QUASAR project. The complex challenge to be addressed is to create conditions under which we can assure equity and access to good mathematics instruction, while simultaneously defining such instruction not in the sense of conventional instruction with its emphasis on memorization, imitation and repetition, but rather in the spirit of the mathematics reform reports, which paint a portrait of school mathematics with textures and hues that emphasize thinking, reasoning, problem solving, and communication.

The Legacy of Conventional Instructional Practice

As many studies (e.g., Porter, 1989; Stodolsky, 1988) have suggested, conventional mathematics instruction emphasizes students learning alone, producing stylized responses to narrowly-prescribed questions for which there is a single answer, which is already known by the teacher, and which can and will be validated only by teacher approval. One consequence of conventional school mathematics instruction can be obtained by inference from the reported tendency of students to perceive school mathematics as a domain which is disconnected from sense making and the world of everyday experience (Resnick, 1988; Schoenfeld, 1991). For example, Silver, Shapiro and Deutsch (1993) found that, when middle school students were asked to provide interpretations for an answer to a division problem dealing with a real world situation, their responses dealt more with technical concerns than with sense making. Many students proposed answers that involved a fraction of a bus (even though they knew that buses do not have fractional parts), apparently because the technical process of computation produced a fractional answer. Students' dissociation of sense making from mathematical activity was evident not only from the responses they provided but also from the explanations they did not give, since reports from the students' teachers suggested that some children engaged in more sense making than was evident in their written responses. Students apparently did not see their "sensible" answers (e.g., using a mini-van as the practical referent for a "fractional part" of a full bus) as having validity as solutions for a school mathematics problem. The requirement that the results of mathematical activity should make sense was apparently not a feature of students' mathematics instruction. Silver et al. also identified another deficiency of conventional mathematics teaching: students had difficulty providing explanations of their reasoning or justifications for their answers. Explanations and interpretations, in oral or written form, are also not a regular feature of instructional activities in mathematics classrooms.

Deficiencies in conventional instructional practice are closely tied to limitations in

curriculum content, especially at the middle school level. Contrary to recent recommendations (e.g., NCTM, 1989) for a broader, enriched curriculum for all middle school students -- a curriculum containing fundamental concepts from statistics, geometry, measurement, probability, and algebra -- conventional mathematics instruction for elementary and middle school students has focused narrowly on the teaching and learning of computational skills involving whole numbers and rational numbers. In a national survey of instructional practices in middle school education, Epstein and MacIver (1989) found a "conservative emphasis on basic skills and a lack of attention to creating learning opportunities that are more responsive to the characteristics of early adolescents" (p. 31). Moreover, in basic subject areas (e.g., mathematics), they found that "the most frequent instructional approaches emphasize drill and practice ... more passive than active learning, and more attention to teaching strategies than to learning strategies" (p. 33). At all educational levels, "drill-to-kill" or "assembly-line" instruction, consisting of repetitive drill and practice on basic computation and other routine procedures, has characterized school mathematics, especially in impoverished urban and rural schools.[1]

Because of the common practice of homogeneous ability grouping, or "tracking", which relegates disproportionate numbers of poor or minority students to a "remedial track" (Oakes, 1990a), instructional deficiencies are often worst for the students who need the most help.

---

[1] This critique of conventional practices in urban and rural schools in poor communities is not meant to deny the tremendous challenges these schools face. Urban and poor schools are more likely to serve populations whose needs are not being met in the areas of health care, housing, transportation, and economic and personal security. As a consequence, poor urban students are less likely than their more affluent suburban counterparts to attend school regularly, to have available energy and attentiveness to focus squarely on an academic agenda, and to be sufficiently free of family and other responsibilities to study well at home. Moreover, inequities in current methods of funding public education in the United States ensure that schools serving students with the greatest needs generally have far fewer dollars to support educational services than schools serving students who live in more economically privileged situations (Kozol, 1991). These conditions make urban education extremely challenging, and their ultimate resolution lies as much in the arena of social and economic policy as it does in instructional practice. Yet, despite these challenges, it is important to note that many urban and rural schools, employing a mixture of courage, tenacity, ingenious leadership, and dedicated staff, have managed to create, at least for a period of time, conditions conducive to good academic education (e.g., Grant, 1988; Meier, 1992).

Academic tracking reinforces students' negative self-perceptions by emphasizing their past failures, and it often has disastrous consequences. Too often, youngsters assigned to lower tracks never recover and find themselves blocked from access to further educational opportunities. The negative consequences of tracking can also be seen in the nature of the instruction students receive in lower-track classes. Data regarding instructional practices suggests that students assigned to the lower tracks of many high schools tend to receive less actual mathematics instruction, less homework, and more drill and practice of low-level factual knowledge and computational skill than students assigned to middle and higher tracks (Oakes, 1985). Although these instructional practices might contribute to increased performance on tasks requiring only basic factual knowledge or on routine computational skills, such practices are clearly unlikely to lead to improvements on more complex tasks requiring mathematical reasoning and problem solving.

If the goal of school mathematics is to help students learn to think and reason about mathematical matters, then instructional activity clearly needs to be quite different from conventional mathematics instruction, especially the instruction received by students who are performing least well in the current system. Knapp and Turnbull (1990) provide an explication of the premises that underlie conventional approaches to teaching "disadvantaged students" and an analysis of the limitations inherent in attempts to use conventional "best practices" to lead to the goals of increased understanding and problem solving in the area of mathematics. They concluded that improvement of the educational situation for students attending schools in economically disadvantaged areas will require more than simply providing greater amounts of the kinds of instruction they now receive. Enhanced forms of mathematics instruction must be made available to all students, including those serving children living in poor communities, and this is an explicit goal of the QUASAR project.

The QUASAR Project: A Brief Overview

In response to the general climate of crisis and need discussed above, the QUASAR project was launched in Fall 1989 as a demonstration that it was both feasible and responsible to implement instructional programs that foster the acquisition of mathematical thinking and reasoning skills by students attending middle schools in economically disadvantaged communities. Arguing that low levels of participation and performance in mathematics for poor urban students were not primarily due to a lack of ability or potential but rather to a set of educational practices that blocked them from meaningful experiences with mathematics learning, QUASAR posited that these students could be assisted to learn a broader range of mathematical content, acquire a deeper understanding of mathematical ideas, and exhibit improved reasoning and complex problem solving, if effort, imagination, and reasonable financial resources were applied.

QUASAR's Instructional Vision

QUASAR rests on the premise that it is both necessary and possible for mathematics education to serve all students well and to provide avenues for them to develop their intellectual potential. Moreover, the project posits that it is possible for such a mathematics education to be consistent with the results of several decades of research on learning, which suggest that learners actively construct their own knowledge, even in complex intellectual domains such as mathematics. The view of learners as active constructors of knowledge suggests the intellectual bankruptcy of previous, deficit-based models of low achievers and suggests a new vision of education. In this view, the task of teachers and schools is not to detect and remediate students' deficits but rather to identify and nurture sources of competence in students. In such an education students would be provided with the necessary support and materials to refine and

make more mathematically sophisticated their own constructs and means of building knowledge, as well as having opportunities to appropriate and use mathematical or general academic concepts, principles, and processes contributed by others. This form of mathematics education is aimed at helping students to use their minds well, rather than teaching them simply to memorize facts and algorithms.

Not only because of the constructivist underpinnings of this vision for mathematics education but also because of the practical demands of providing rich learning opportunities for diverse populations of children, QUASAR asserts that increased pedagogical emphasis must be placed on assisting learners to engage in mathematical activity which is embedded in the learner's social and cultural context. It is essential for instruction to address the connection between the mathematics taught in school and the social lives of the children who are asked to learn it. Thus, educational practices must embrace, affirm, and begin with the content and structure of what students bring to the enterprise. As Ernest (1991) has argued:

> School mathematical knowledge must reflect the nature of mathematics as a social construction: tentative, growing by means of human creation and decision-making, and connected with other realms of knowledge, culture and social life. ... it is to be embedded in student culture and the reality of their situation, engaging them and enabling them to appropriate it for themselves (pp. 207-208).

QUASAR seeks a new form of high-literacy education that blends attention to basic-level and high-level mathematical goals and produces students who not only can accurately execute algorithms and recall factual knowledge but also have the capacity to impose meaning and structure on new situations, to generate hypotheses and critically examine evidence, and to select the most appropriate from among a repertoire of strategic alternatives. In such an education, students would not only learn to read, write and perform basic arithmetic procedures, but also

learn when and why to apply those procedures, learn to make sense out of complicated situations, and learn to develop strategies for formulating and then solving complex problems.

This vision of mathematics education places social interaction and communication at the heart of meaningful learning. Mathematics classrooms must become communities of collaborative, reflective practice, in which students are challenged to think deeply about and to participate actively in engaging the mathematics they are learning. As Silver, Kilpatrick and Schlesinger (1990) have argued, "Within communities, the need for communication is obvious. Within mathematical communities, communication in the form of discussion, argument, proof, and justification is natural" (p. 23). In these classroom communities, students would be expected not only to listen but also to speak mathematics themselves, as they discuss observations and share explanations, verifications, reasons, and generalizations. In such classrooms, students would have opportunities to see, hear, debate, and evaluate mathematical explanations and justifications, because "the emphasis is less on memorizing procedures and producing answers and more on analyzing, reasoning and becoming convinced" (Silver et al., 1990, p.38). This view of mathematics classrooms is compatible with the findings of Resnick (1987), who reviewed research on teaching high-level thinking and reasoning skills and concluded that developing higher-order cognitive abilities requires shaping a disposition to thought through participation in a social communities that value thinking and independent judgment. Thus, such classroom communities represent a new vision of mathematics education -- a vision compatible with the precepts of the contemporary reform documents and aimed at eradicating the legacy of conventional instructional practices and allowing equitable access and for all students to high quality mathematics instruction and challenging content.

Beyond their value in providing opportunities for more authentic forms of mathematical activity and student discourse (NCTM, 1991), such mathematics classrooms have design features

that make them highly likely to be supportive of the learning of culturally diverse students. A recent examination of educational practices used with linguistically and culturally diverse student populations found that collaboration and communication were key elements of effective instructional practice at all educational levels and that the curriculum in successful programs contained a blend of both challenging and basic academic material (Garcia, 1991). Thus, it is reasonable to promote the development and implementation of this form of instruction for all students, especially if such instruction can also be attentive to the needs, interests and backgrounds of culturally diverse students. Further details concerning this instructional vision and the way it is realized in QUASAR classrooms are provided by Silver, Smith and Nelson (in press).

Some Design Principles and Features of the QUASAR Project

The QUASAR reform strategy combines elements of "top down" and "bottom up" approaches to school change. In the tradition of "top down" reform efforts, the importance of coherent general principles as guides for reform is recognized and all project sites have affirmed the general goals of curriculum breadth, deeper student understanding, and emphasis on high-level thinking and reasoning; local project sites have also developed plans that incorporate a shared set of focal activities: staff development, ongoing teacher support, curriculum development or revision, and alignment of student assessment with instructional practice. On the other hand, recognizing the power of "bottom up" approaches to reform and the importance of tying reform efforts closely to the nuances of local conditions, QUASAR does not encourage or support reform imposed from a distance. Rather, the project encourages and supports reform efforts that are designed and implemented by those who live or work in the affected communities. By working with locally-based collaborative teams, the strengths of each member

of the partnership can be utilized and the programs can be woven into the educational and social fabric of the schools and surrounding communities, in order to build the capacity of those schools and communities to face fundamental challenges and to solve their own educational problems.

QUASAR is not only a practical school demonstration project; it is also a complex research study of educational change and improvement. The project's research design has been heavily influenced by evidence -- accumulated from several decades of research on school reform -- that school change must be treated as a *process* rather than as a *product* (e.g., Lieberman, 1986). Therefore, the project seeks to study programmatic activities as they occur rather than waiting until an appropriate moment in time to render a summative judgment about the effectiveness of the reform efforts. Project research aims to identify critical features of successful programs by studying several different approaches being taken to accomplish the general instructional program goals; examining the implementation of these programs in schools and in teachers' classrooms; assessing the impact of the programs on teachers' instructional practices, knowledge and beliefs; evaluating the impact of the programs on student performance by devising new assessment tools to measure students' growth in mathematical reasoning and problem solving; and ascertaining conditions that appear to facilitate or inhibit the success of these instructional reform efforts. Through its extensive research effort, the project aims to identify instructional programs, practices, and principles that can guide effective mathematics instruction for middle school students and to describe key features of good instructional programs so that they can be adapted to other schools.[2]

---

[2] Many research aspects of the QUASAR project are not discussed in this paper. Descriptions of the research design and methodology being employed to examine teachers' instructional practices can be found in Stein, Grover, and Silver (1991a, 1991b). It is worth noting that there are important equity-related issues embedded in the research aspects of the project. For example, efforts are made to ensure racial and ethnic balance among the observers who document classroom activity at the project sites, and students who are identified for special observation in classrooms are drawn from a pool that is balanced for gender, ability and ethnicity. Although these

QUASAR Sites and Programs

School sites and their surrounding communities constitute the operational heart of project activities. QUASAR has begun its work with a small number of educational partnerships centered around middle schools located in economically disadvantaged areas. In particular, six geographically dispersed sites are serving as initial development environments for teachers and administrators from a middle school, working in collaboration with "resource partners" from a local university or education agency, to develop, implement, and modify innovative mathematics instructional programs for middle school students. Across the six sites there is considerable diversity in the ethnicity and race of the student populations, with two sites serving predominantly African-American students, two serving primarily Hispanic-American students, and the other two sites having more culturally diverse student populations.

In line with the general goals of the project, the mathematics curriculum at these sites is being broadened to include treatment of a wide array of mathematical topics that stretch beyond computation with whole numbers and fractions, and the content and instructional practices are being enriched through an emphasis on thinking, reasoning, problem solving, and communication. Key features of instruction in most classrooms in QUASAR schools include student engagement with challenging mathematical tasks, enhanced levels of student discourse about mathematical ideas, and student involvement in collaborative mathematical activity. In addition, teachers and resource partners seek ways to connect the content and form of mathematics instruction more closely to children's natural ways of thinking and reasoning and to their lives and experiences outside of school (e.g., everyday problem settings, culturally relevant teaching activities). For a detailed discussion of QUASAR instructional programs in

---

research-related equity issues are not discussed in this paper, they are being addressed in the QUASAR project.

relation to concerns of equity and mathematical quality, see Silver, Smith and Nelson (in press).

In recognition of the complexity of the project's goals, a broad array of activities are undertaken at project sites, including curriculum development and modification, staff development and ongoing teacher support, classroom and school-based assessment design, outreach to parents and the school district at large. It is beyond the scope of this paper to describe these efforts in detail. Suffice it to say that instructional improvement efforts are being supported by a network of interrelated activities that attempt to develop the capacity of the school and the teachers to provide an enhanced mathematics program for each child. For example, in the area of teacher staff development and ongoing support, QUASAR sites are characterized by a diverse set of activities, including regular meetings at which teachers can discuss instructional goals and share the results of their implementation efforts; regular interactions with the resource partner(s); specially designed courses or formal staff development sessions on topics of interest to the teachers; "retreats" to provide time for reflection and extended discussion of progress; and participation in professional meetings and conferences.

QUASAR teachers are ordinary people engaged in extraordinary efforts, under very difficult circumstances, to develop enhanced instructional programs for their students. Collectively, the collaborators at project sites are engaged in efforts that reside at intersection of mathematics educational reform and attempts to increase educational equity. In their programmatic work they are attempting to increase students' access to strong, non-remedial curriculum and instruction; to increase students' confidence and competence in using mathematics to solve problems; to increase students' interest in mathematics; to increase students' understanding of the mathematics they learn; and to increase students' ability to communicate about the mathematics they learn, especially with respect to reasoning and constructing mathematical arguments. QUASAR teachers and resource partners are engaged in

the hard work of bringing the rhetoric of mathematics education reform into contact with the realities of urban schooling.

The Nature and Role of Assessment in QUASAR

Assessment activities in QUASAR are guided and constrained by their being embedded within the larger project. Given the somewhat experimental nature of the instructional activity at the various sites, student assessment in QUASAR has been viewed, in large part, as a measure of instructional program accountability. Thus, student assessment data are viewed as important information concerning the extent to which students reap the intended benefits from their school mathematics experiences. In this regard, it was deemed sufficient that the obtained information reflect programmatic outcomes for groups of students rather than providing individual student-level reports. Given the nature of the project as a longitudinal demonstration that appropriately enhanced mathematics instruction could be of substantial benefit to students, it was viewed as essential that the assessment instruments be technically sound, thereby facilitating the production of credible, convincing evidence regarding instructional efficacy.

Given QUASAR's promotion of mathematics instructional goals compatible with new national standards for mathematics proficiency, it is essential that its assessment instruments reflect these goals (e.g., reasoning, problem solving, communication). At the time the project was designed and launched in 1988-89, there were prototypes of some types of tasks that might be used in such an assessment, but there were no existing instruments for middle school mathematics that had sufficient reliability and validity evidence to support their use and score interpretation. Most assessment instruments for grades 6-8 were standardized, multiple-choice tests that required rapid responses to questions which tapped primarily procedural facility and factual knowledge but which provided few opportunities for students to demonstrate

mathematical reasoning, problem solving and communication. Thus, in order to ensure that appropriate measures were available to monitor and evaluate program impact, it was necessary for QUASAR to develop its own assessment instruments.

The QUASAR project employs a variety of measures in assessing student growth, including paper-and-pencil cognitive assessment tasks administered to individual students in a large group setting; "instructionally-embedded" tasks administered to students in natural classroom settings, such as cooperative learning groups, and on which students are able to work collaboratively; individually administered performance assessment tasks, which may involve the use of manipulative materials and computational tools; and non-cognitive assessments aimed at important attitudes, beliefs, and dispositions. In addition, teachers at the project sites also supply information available from their own classroom sources (e.g., tests, homework, projects) and administrators provide information regarding performance on district-mandated tests to supplement the store of information about both the program and individual students.

In the next section of this paper, details concerning one major component of assessment activities within the project -- the QUASAR Cognitive Assessment Instrument (QCAI) -- are presented. In particular, information is provided concerning the design principles for development and validation of the assessment instrument and associated guides for scoring student responses, the administration procedures, and the nature of information reported to teachers and administrators about student performance. Issues related to mathematics content appropriateness, technical measurement quality, and equity issues are interwoven throughout.

The QUASAR Cognitive Assessment Instrument: Development, Administration, and Validation

The QCAI is designed to measure student outcomes and growth in mathematics, and to help evaluate attainment of the goals of the mathematical instructional programs (Lane, 1993;

Silver & Lane, 1993). The QCAI assesses student performance on open-ended tasks involving mathematical problem solving, reasoning, and communication. Throughout the development phase steps are taken to ensure that the QCAI reflects current understandings about mathematical problem solving and reasoning and the acquisition and use of mathematical knowledge and skills, and that it reflects the contemporary view of mathematical proficiency, with its emphasis on reasoning, problem solving and communication (NCTM, 1989). As was discussed earlier in this paper, this view of mathematical proficiency, as well as the prevailing assumption that all students should be expected to acquire such proficiency, is compatible with the mathematics curriculum and instruction at the participating schools. Thus, a major consideration in the development and administration of the QCAI is ensuring that it provides a valid assessment of all students mathematical thinking and reasoning, and their acquisition of knowledge about a broad range of mathematical topics.

In designing and revising the QCAI a number of factors are considered. These factors include ensuring the quality of the assessment from the perspectives of mathematical content quality, psychometric technical quality, and equity and fairness. Also considered are factors related to practical constraints, such as the amount of time available for administration. For a particular grade level, the QCAI[3] consists of 36 tasks which are distributed into four booklets, each containing nine tasks (Lane, Stone, Ankenmann, & Liu, 1992). The four booklets are randomly assigned to students within each classroom. Thus, each student receives only one booklet that is to be completed within one class period (approximately 40-45 minutes). The use of a matrix sampling approach allows for the assessment of students in a relatively short time frame, thereby keeping interruptions to the instructional process minimal; avoids the problems

[3] The same version of the QCAI is used for both 6th and 7th grade students, whereas, the version of the QCAI for the 8th grade students consists of some tasks that are in the 6th/7th grade version as well as tasks that are unique to the 8th grade version.

associated with sampling only a small number of tasks (Mehrens, 1992); and affords valid generalizations about students' mathematical proficiency at the program level.[4] The QCAI is administered in the fall and spring of each instructional year, and each student receives a different booklet of tasks at each administration occasion, thereby avoiding the problems associated with using the same tasks to assess students over time. Because of these design features, confidence is increased that changes in student performance can be contributed to increased mathematical proficiency rather than to prior knowledge of the assessment tasks. It should be noted that the content of the QCAI is modified somewhat each year, in order to allow release of some tasks and to broaden the range of content and processes assessed.

This section of the paper describes the process used for developing, administering, and validating the QCAI, in particular, it highlights the interactive nature of the development, administration, and validation process. Special attention is given to describing how validity evidence is obtained for content quality and representativeness, cognitive complexity, curricular relevancy, consequences of the use of the assessment and interpretation of the scores, and fairness of the assessment (Dunbar, Koretz, & Hoover; 1991; Linn, Baker, & Dunbar, 1991; Messick, 1989). As described in the Standards for Educational and Psychological Testing, validity "refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation is the process of accumulating evidence to support such inferences (American Educational Research Association et al., 1985, p.9)." For example, if the intent of an instrument is to assess complex cognitive skills and inferences from the test scores reflect this intent, both logical and empirical evidence needs to be collected to support such inferences.

---

[4]Technical information on the reliability and validity of the QCAI, in particular, the degree to which one can generalize from the scores derived from the assessment to the broader construct domain of mathematics is provided in Lane, Stone, Ankenmann, and Liu (in press).

The fairness of an assessment is closely connected with all sources of validity evidence. As described by Cole and Moss (1989), "bias is differential validity of a given interpretation of a test score for any definable, relevant subgroup of test takers (p. 205)." This conception of bias as differential validity suggests that ensuring a fair assessment requires evidence to support the appropriateness, meaningfulness, and usefulness of the inferences made from test scores for all students. As an example, for an assessment to be valid it needs to be capable of evoking the same level of cognitive activity for all groups of students regardless of their gender, cultural, ethnic, or linguistic backgrounds. The wording and context of the tasks, for example, should not interfere differentially with student performance.

Validity Considerations in the Design of the QCAI

Assessments that allow students to display their thinking, reasoning, and strategic processes that underlie their performance can ensure more valid inferences regarding the nature and level of students' understanding (Lane & Glaser, in press). Open-ended assessment tasks that ask students to show their solution strategies and provide justifications for their answers or conjectures allow for the display of various levels of student understanding. Consequently, such tasks have the potential to be sensitive in measuring changes in student achievement. Thus, open-ended tasks are used in the QCAI.

An instrument consisting of open-ended tasks is capable of assessing students' knowledge and achievement in complex construct domains. This implies, however, that the breadth of the assessment instrument needs to reflect the breadth of the complex construct domain underlying the interpretation of the scores (Cronbach & Meehl, 1955; Messick, 1989). Thus, to ensure valid generalizations from the scores derived from the assessment, the specifications used for the QCAI need to reflect the breadth of the complex construct domain of mathematics.

Valid assessment practices are linked to the procedures adopted in the development and administration of the assessment instrument. Thus, a valid assessment of all students, regardless of their cultural, ethnic, and linguistic backgrounds, requires that the development, administration, and validation of the assessment instrument be interwoven (American Educational Research Association et al., 1985; Duran, 1989). Furthermore, the development and validation of an assessment is an ongoing activity. This implies that validity evidence for the use of the QCAI with students from various cultural, ethnic, and linguistic backgrounds needs to be collected continuously and systematically as the instrument is being developed, administered, refined, and extended.

With respect to national standardized-testing situations, Pollack, Rock, and Jenkins (1992) comment that "the test developers, test takers, and scorers are strangers to each other" (p. 6). In stark contrast, within the QUASAR project test developers have a close relationship with the participating schools and as a consequence, we are continuously acquiring knowledge about the nature and goals of the instructional programs and about the ethnic and cultural backgrounds of the students at the participating schools. In addition, the instructional programs at the schools are similar in the sense that they have an overarching common goal - to provide ample opportunities for all students to think, reason, and communicate mathematically - although they may take different, but not incompatible, avenues to reach the same goals. Such familiarity with the participating schools, and the fact that the schools are striving for the same outcomes, helps ensure that the QCAI reflects the common themes across the instructional programs and is sensitive to the nature of the students. For example, in developing and administering the QCAI procedures are used to minimize the measurement of irrelevant (or incidental) constructs. As Messick (1989) indicates, if irrelevant constructs are being assessed in addition to the construct of interest, an assessment may be more difficult for some groups of students, thereby resulting

in scores that are invalidly low for the adversely affected groups. Thus, in designing the QCAI, consideration is given to the amount of reading and writing required in responding to tasks, as well as the likely familiarity of the task contexts for culturally diverse students. QUASAR's knowledge of the student population to be tested allows the embedding of QCAI tasks in reasonable and appropriate contexts.

Steps are also taken to help ensure that the assessment will not produce scores that are invalidly high for some groups of students. For example, if some students are more familiar with the task formats than others, they will be at an advantage in responding to the assessment. For all students to have the same opportunity in displaying their reasoning and thinking they need to understand the nature of the assessment tasks as well as the nature of expected performance. Our knowledge of the instructional programs and some of the instructional activities allows us to be sensitive to the unique characteristics of each school and to the common characteristics across the schools (e.g., formats and directions used in classroom tasks) when designing the QCAI tasks.

Specification of the Assessment Tasks

As noted earlier in this paper, current conceptualizations of mathematical proficiency emphasize understanding and applying mathematical concepts, principles, and procedures; discerning mathematical relations; making connections among mathematical topics and between mathematics and the world outside the mathematics classroom; solving complex mathematical problems; reasoning mathematically; and communicating mathematical ideas (NCTM, 1989). In this view, mathematics is seen to involve problems that are complex, yield multiple solutions, require interpretation and judgment, require finding structure, and require finding a solution path that may not be immediately visible (NRC, 1989). This conceptualization of mathematical

proficiency is reflected in the specifications used for developing the QCAI and is in alignment with the view of mathematics that is guiding the development and implementation of instructional programs at QUASAR sites.

The development and review of QCAI tasks and scoring rubrics involves mathematics educators, mathematicians, cognitive psychologists, psychometricians, and multicultural educators, thereby ensuring that the specification of the QCAI blends considerations of mathematical content quality, current conceptualizations of mathematical proficiency, contemporary perspectives on student learning and understanding, as well as important equity and psychometric issues. Table 1 shows the specifications for the QCAI which include four major components: mathematical content, cognitive processes, mode of representation, and task context.

Insert Table 1

To some extent, the components, and the categories within the components, are interrelated; therefore, the framework is conceptualized not as a matrix with discrete cells but as seamless fabric. Such a conceptualization allows for an individual task to assess topics in more than one content area and to assess a variety of processes. This facilitates the development of tasks that assess complex mathematical thinking. For example, a task may assess students' understanding of statistical concepts and ability to reason proportionally as well as ability to organize information, use appropriate strategies and procedures, and generalize results. Such a task may be embedded in a real world context and be represented in both text and a table. It may also allow students to display their thinking and knowledge using a variety of representations.

A number of task formats are used to ensure that the complexity of the domain of

mathematics is captured by the assessment (Lane, Parke & Moskal, 1992). For example, some of the tasks ask students to provide a justification for their answers while others ask students to show how they found their answers or provide a description of presented data. Figure 1 provides some examples of sample tasks and desired student responses.

Insert Figure 1

Because the assessment includes a number of task formats a variety of representations, strategies, and processes can be elicited from the students. By allowing for a variety of representations (e.g., written, pictorial, numerical), students who are not proficient in writing are less likely to be at a disadvantage because they may use pictorial and numerical representations to display their understanding in addition to written text. Furthermore, through a process of self-documentation, QUASAR sites provide to project staff samples of student responses to classroom instructional and assessment tasks. This information not only illuminates the content and processes being assessed in the classrooms and the evedient level of student understanding, but it also illustrates important aspects of task presentation and directions. This information is then used to guide QCAI task development to ensure reasonableness for students at the various schools.

Review of the Tasks and Student Responses

Internal review. Assessments need to be appraised with regard to the quality and comprehensiveness of the content and processes being assessed and with regard to bias in task language and context. Both internal and external reviews of the assessment tasks are conducted (Lane, 1993; Lane, Parke, & Moskal, 1992). The internal review is an iterative

process in that when a task is developed it may be reviewed and modified a number of times prior to and after being piloted. This involves a logical analysis of the task to ensure it is assessing important content and processes, worded clearly and concisely, and free from anticipated sources of bias. Some of the questions that are addressed to help ensure that the tasks are free from anticipated sources of bias are: Is the task context likely to be familiar to all groups of students?, Are different ethnic and cultural groups represented in the tasks favorably?, and Are the tasks clearly worded to ensure that all students understand what is expected? In addition, members of the project who work closely with the participating schools and are familiar with the instructional programs review the tasks to help determine whether the tasks are reflective of the goals of the instructional programs at the schools. It should be noted that some of the tasks are discarded prior to reaching the pilot phase.

The data from the pilot provides evidence indicating whether the tasks are assessing the content and processes that they were intended to assess and whether the wording and directions of the tasks are interfering with student performance (Lane, Parke, & Moskal, 1992). As Messick has indicated, a variety of methods can be used to analyze the processes underlying task performance (Messick, 1989). For example, students can be asked to think aloud as they solve problems or provide a verbal or written rationale for their responses. Inferences about students' thinking can also be made from an analysis of their errors. These methods were used in the individual and group pilot administrations of the assessment tasks. The tasks were piloted with students from the participating schools and with students who had similar backgrounds to the students at the participating schools.

In the individual administrations students are asked to think aloud as they solve the problems. This affords rich information from a relatively small number of students regarding the degree to which the tasks evoke the content knowledge and complex processes

that they were intended to evoke, and allows for additional probing regarding the processes underlying student performance. The individual administrations also provide an opportunity for the examiner to pose questions to students regarding their understanding of task wording and directions. The group administrations provide a large number of student paper and pencil responses that are analyzed to ensure that the tasks evoke the content knowledge and cognitive processes that they were intended to evoke, the directions and wording are as clear and simple as possible, and misconceptions in students' thinking can be detected from their written responses. Multiple variants of a problem are piloted to further examine the best way to phrase a problem to ensure that all students have the same opportunity to display their reasoning and thinking. Information from the pilot analyses is used in the task revision process.

External review. Mathematics educators, psychometricians, and cognitive psychologists from various cultural, linguistic, and ethnic backgrounds review the tasks to ensure that they measure the content and processes that were intended to measure, the task wording is as short as possible and free of unnecessary verbiage, and the task directions indicate clearly what is expected of the students. The tasks are also reviewed for their perceived fairness. The reviewers are asked to indicate whether the task context, wording, and format would be familiar and mean the same thing to students with various cultural, linguistic, and ethnic backgrounds. The external reviewers are also asked to indicate whether the assessment as a whole represents the construct domain of mathematics. The suggestions and comments made by the external reviewers are considered in the process of refining the tasks.

Specification of Scoring Rubrics

A focused holistic scoring procedure was adopted for the scoring of student responses. This was accomplished by first developing a general scoring rubric that reflected the conceptual framework that was used for constructing the assessment tasks. The general rubric incorporates three interrelated components: Mathematical conceptual and procedural knowledge, strategic knowledge, and communication (see Appendix A). It should be noted that the rubrics developed by the California Assessment Program (California State Department of Education, 1989) for their open-ended assessment tasks contributed in the conceptualization of the QCAI general scoring rubric. In developing the general scoring rubric, criteria representing the three interrelated components were specified for each of the five score levels (0-4) (see Appendix B). Five score levels were used to facilitate in capturing various levels of student understanding.

Based on the specified criteria at each score level, a specific rubric is developed for each task. The criteria specified at each score level for each specific rubric is guided by theoretical views on the acquisition of mathematical knowledge and processes assessed by the task, and the examination of actual student responses to the task. The examination of the student responses from the participating schools helps ensure that the rubrics reflect the various representations, strategies, and ways of thinking that are common across the schools as well as those that are unique to one or more schools. The process for developing a specific scoring rubric is iterative in the same way that the process for developing a task is iterative. Furthermore, the scoring rubrics are subjected to a review process which is similar to the review process used for the assessment tasks.

Promotion of Awareness of QCAI Expectations

To ensure that students are familiar with QCAI task formats and the criteria used to

evaluate performance, prior to the fall administration each year, teachers are sent sample tasks, examples of scored student responses, and criteria for assigning scores . Teachers can then use the sample tasks with their students and discuss what the tasks are assessing and the criteria used for scoring student performance. Not only does this practice help ensure an equitable distribution of task familiarity across different sites but it also contributes to the systemic validity of the assessment, since transparency -- access to the criteria for evaluating performance -- is critical to encouraging systemic validity (Frederiksen & Collins, 1989). Teachers are also encouraged to involve their students in this activity prior to the administration of the QCAI in the spring of each year as well as throughout the instructional year not only with the QCAI sample tasks but also with their own tasks.

An interesting example of the systemic or consequential validity of the QCAI as part of the QUASAR project is embedded in the experience of the teachers at one of the project sites during the second year of the project using the QCAI practice tasks. One of the tasks that was sent and administered by the teachers was the Busy Bus Company Problem (shown in Figure 1). The teachers followed the administration guidelines for the tasks, but did not initially refer to the other information sent with the tasks, such as suggestions for evaluating students' responses. After administering the QCAI practice tasks, teachers met to discuss their students' performance, and some surprising findings regarding the Busy Bus Problem were revealed. In particular, teachers reported that many students indicated that Yvonne should purchase the weekly pass rather than paying the daily fare, which the teachers believed to be the more economical choice. Curious about this unexpected answer to what the teachers believed to be a rather straightforward question -- a multi-step arithmetic story problem involving multiplication of whole numbers -- they decided to discuss the problem in class and ask students to explain their thinking.

The ensuing discussion with students provided an interesting illustration of students applying out-of-school knowledge and problem-solving strategies to a mathematics problem. Many students argued that purchasing the weekly pass was a much better decision because the pass could allow many members of a family to use it (e.g., after work and in the evenings), and it could also be used by a family member on weekends. Students' reasoning about this problem -- situated in the context of urban living and the cost-effective use of public transportation -- demonstrated to the teachers that there was more than one "correct" answer to this problem, which was the intent of the task developers, as can be seen by the description of desirable responses in Figure 1. But it was also clear to the teachers that, although many students were applying sound real-world-based reasoning to the solution of this problem, too few had sufficiently connected their reasoning to the relevant mathematics, as is done in the sample desirable response in Figure 1, in order to provide a response that would be judged to be of high quality. Thus, teachers used this discussion as an occasion to illustrate the criteria for a complete, high quality explanation that was based both on solid mathematics and on sensible reasoning. As was noted earlier in this paper, the linkage of school mathematics to sense making is a goal of the QUASAR project, since conventional mathematics instruction has tended to divorce mathematics from sense making.

This experience made it clear to the teachers that if their goal was assessing what students know and are able to do, then it was essential that students not only provide answers but also explain their thinking and reasoning. Moreover, it was clear that students needed assistance in understanding the criteria for strong and weak mathematics responses. The subsequent assessment practices of the teachers at this school reflected this understanding and has led to a dramatic improvement in the quality of their locally-developed assessments. In fact, their school-based assessment work has drawn favorable attention from district

administrators seeking ways to implement performance assessments on a broad basis in the school district. Thus, the teachers' experience with QCAI tasks influenced programmatic developments compatible with the goals of increased student access to high-level mathematics.

Design and Administration Considerations for QCAI Bilingual Versions

As indicated by Duran (1989), when assessing language minority students, an assessment instrument needs to be appropriately translated and language interpreters should help in its administration. Furthermore, research has indicated that the translation needs to be specific to the intended target student population because target groups can vary considerably in their familiarity with dialects and varieties of English and non-English languages, and familiarity with the task wording can affect student performance (Duran, 1989). Spanish bilingual versions of the QCAI are developed in which both the English and Spanish version of the task is presented in the same booklet and students have the option to read the task in Spanish and/or English and to respond in either language. Because the participating schools are in different locations in the country, some of the Spanish speaking students have a Mexican American heritage while others have a Puerto Rican heritage. Consequently, if a word or phrase in a task is different in the two Spanish dialects both versions are included. To ensure that the Spanish version is parallel to the English version it is backtranslated so that the task developers can examine it in relation to the English version, and if necessary revisions are made to the Spanish version. In addition, the student responses form the pilot are evaluated to ensure that the tasks evoke similar thinking on the part of the students regardless if they received a bilingual version or an English version. It should be noted that our conception of language minority students is broad, in that, it may

include not only individuals that have Spanish, for example, as their primary or secondary language but also individuals who speak different dialects of English. As previously indicated, the wording and directions of the tasks are reviewed internally and externally to help ensure that students who may speak different dialects of English understand the problem situation and what is expected.

When administering the QCAI, all students are encouraged to ask the examiner for help if they have difficulty in reading a word or phrase. This helps ensure that the ability to read either English or Spanish does not interfere with performance since research has indicated differences between verbal and nonverbal assessment performance that implicate language proficiency as a factor affecting performance (Duran, 1989).

Ascertaining Alignment of the QCAI with the Instructional Programs

In the spring of each instructional year, teachers are asked to provide information on the degree to which the mathematical content and processes assessed by the tasks are consistent with the goals of their instructional programs. For each task, teachers are provided with a description of the task and are asked to indicate whether the task content and processes have been included in their instruction for each of their mathematics classes. This is done for equity reasons -- to ensure that the results of the QCAI can be interpreted fairly with respect to nuances of individual programs -- and as a measure of the content quality of the instructional programs, since the QCAI contains a reasonable coverage of expectations for middle school mathematics instruction. Given that some of the QCAI tasks cover several grade levels, the extent of alignment reported by teachers at each grade level has been acceptably high. For example, in the 1990-1991 instructional year, in an average of 73% of the 6th grade classes, teachers indicated that the content and processes assessed by the QCAI

tasks had been taught.

Analysis of Student Responses

Student responses are rated by middle-school mathematics teachers. The raters score the student responses after they are formally trained. First, the general rubric is presented and discussed. Then the specific rubric for a task and prescored student responses are presented and discussed. The raters then practice scoring student responses, and their scores are discussed in relation to the scores previously assigned by a member of the assessment team. Finally, the raters score the actual student responses. Each response is scored independently by two raters. If the raters disagree by more than one score level, a member of the assessment team also rates the student's response. The Spanish responses are scored by individuals who are fluent in Spanish and who have also had the formal training. It should be noted that information that could potentially bias the raters' assigned scores (e.g., student and school names) is removed from the assessment booklets. A discussion of additional considerations in designing QCAI rating sessions is provided by Kenney and Tang (1992).

In addition to scoring the student responses as described above, a small number of the tasks, which are then "released" and no longer used in subsequent assessments, are evaluated using another procedure. This analysis provides information on various strategies and representations used by students, common misconceptions displayed in student responses, and the degree to which students can communicate their mathematical knowledge.

Reporting of the Results

Reports based on each year's fall and spring administrations of the QCAI are developed and sent to the schools. One report provides information, for each grade level, on

the percentage of students responding at each of the 5 score levels for each of the tasks. For example, one of the three tasks that were released in the summer of 1992 was the Block task which is presented in Figure 2. This task assesses a student's understanding of number sense and problem solving ability using basic concepts of number theory. The student needs to solve for an unknown number which satisfies several conditions set in the story context. Several problem solving strategies can be used such as finding common multiples and interpreting remainders in division computations. Students may also use pictures, words, or mathematical expressions to represent their solutions. To receive a score of 4, a student's work would need to show a correct and complete understanding of the conditions of the problem and of common multiples; and to receive a score of 3, the work would need to be esentially correct, except for a minor error. To illustrate the difficulty of the task and the possibility of its measuring improved performance over time, consider that only 9% of the 6th grade students scored either a 3 or 4 on this task in Fall 1990; whereas, 22% of those students scored either a 3 or 4 in Spring 1992 when they were students in 7th grade. Figure 2 shows examples of students' responses in each of the five score levels.

Insert Figure 2

A more detailed report is also provided for a few of the QCAI tasks, which are then "released" and no longer used in subsequent assessments. This report provides the qualitative analysis of student responses, focusing on examples of differing levels of student understanding, various strategies and representations used by students, and common misconceptions displayed in students' responses. Teachers are also provided with the actual responses to these tasks made by students in their school and encouraged to evaluate the

student responses using our criteria or locally-invented criteria. They are also encouraged to discuss the tasks and evaluation criteria with their students as part of an ongoing effort to increase students' understanding of the nature of mathematical proficiency. These reports also serve as the basis for discussions among teachers and resource partners in developing plans for program improvement.

The use of QCAI information and reports in this way provides another example of the systemic validity of the assessment as part of the QUASAR project. As Frederiksen and Collins (1989) indicate, "A systemically valid test is one that induces in the educational system curricular and instructional changes that foster the development of the cognitive skills that the test is designed to measure" (p. 27). Evidence for the systemic validity of an assessment is provided if it encourages behaviors on the part of teachers and students that promote the learning of valuable skills and knowledge (Frederiksen & Collins, 1989; Linn, Baker, & Dunbar, 1991; Messick, 1989). The QCAI qualitative reports provide information that can be used by teachers and resource partners to improve the instructional programs. By examining the nature of students' responses and the kinds of strategies and errors that were made by students, teachers can plan appropriate instructional modifications or enhancements. In addition, the analyses provided in these reports can provide evidence of the mathematical content quality and cognitive complexity of the QCAI, as has been demonstrated by Magone, Cai, Silver, and Wang (in press) for three tasks that were released after the first year.

Conclusion

This paper has discussed the design, implementation, and validation of the QUASAR Cognitive Assessment Instrument, which is a mathematics assessment instrument comprised of open-ended tasks. The QCAI has been discussed with respect to its function within the

QUASAR project, which is aimed at improving the quality of mathematics instruction for students attending middle schools in economically disadvantaged communities. The discussion has identified many of the ways in which the development, implementation and validation of the assessment has attended to balancing considerations of mathematical content quality, technical measurement issues, and equity concerns.

Within the project, the QCAI provides important data regarding the extent to which students reap the intended cognitive benefits from the instructional programs in QUASAR schools. The QCAI serves as a measure of instructional program accountability, and it exists in a dynamic relationship with other aspects of the QUASAR project. Because the QCAI is embedded in a larger effort to improve instructional program quality, its results are presented in a variety of ways to promote instructional program improvement.

Focusing on the QCAI as an assessment instrument *qua* assessment instrument, evidence for its validity was extensively discussed. In order to have assessment results that provide credible evidence of instructional program impact on students, potential sources of invalidity are minimized in the process of developing and administering the QCAI. The efforts that have been described herein afford evidence of the validity of the QCAI for culturally and linguistically diverse students. It is also expected that this validity evidence will be enhanced over time by additional analyses. For example, the validity of the assessment for students whose native language is Spanish is being examined through individual interviews that probe students' comfort with reading and then responding to either English or Spanish language versions of QCAI tasks. Also, the tendency of bilingual students to provide written responses in either language to tasks presented in both languages is being examined.

These analyses, and others like them, should allow examination of the possibility of

differential task performance by groups of similarly capable students from different racial, ethnic or linguistic subgroups. Within QUASAR , such analyses are preferable to more traditional methods (i.e., studies to detect differential item functioning) because the distribution of students at project sites confounds the demographics of race, ethnicity and language with instructional program differences (i.e., there are possibly important differences in the instructional programs at project sites, yet the student population at two sites is primarily African-American, and at another two sites the student population is primarily drawn from populations whose native language is Spanish rather than English). These analyses will supplement other studies that evaluate the extent to which differential performance on individual QCAI tasks may be due to differences in the instructional programs across sites. This collection of studies will provide information that will illumine not only the instructional impact that may be due to unique aspects of each of the instructional programs or to common program elements but also the possibility that different groups of students within the culturally and linguistically diverse student populations may be impacted in different ways.

Validation of an assessment involves not only accumulating logical and empirical evidence to support inferences made from the scores but also evaluating the consequences of using the assessment instrument (Messick, 1989). In fact, it has argued that "if performance-based assessments are going to have a chance of realizing the potential that the major proponents in the movement hope for, it will be essential that the consequential basis for validity be given much greater prominence among the criteria that are used for judging assessments" (Linn et al., 1989, p. 17). The QCAI provides evidence of the impact of instruction on students, and the QCAI results can be used in turn by teachers and resource partners at project sites to improve instruction, thereby providing evidence of the

consequential validity of the QCAI. Evidence of these consequences can be seen in other data collected by the project, such as in direct observations of classroom instruction or in interviews that are regularly conducted with teachers and resource partners. Evidence of possible consequences of the QCAI can be seen in increased classroom use of open-ended tasks that assess complex mathematical thinking; in the use of complex, QCAI-like scoring criteria for school-based, locally-developed assessments; and in improved student performance over time on the QCAI. Naturally, none of these is clear evidence of the impact of the QCAI alone, since there are many convergent forces acting to shape the instructional programs in desirable directions and to improve student performance, but that is exactly as it should be. As Messick (1992) has asserted, "...in practice, the issue may not be just the systemic validity of the tests but, rather, the validity of the system as a whole for improving teaching and learning (p. 15)."

Because the QCAI and the instructional programs in QUASAR schools share the common, prevailing assumption that all students can acquire knowledge and skills in a broad range of mathematical topics and develop increased proficiency in the areas of mathematical reasoning, problem solving, and communication, the accumulation of evidence for the consequential (or systemic) validity of the QCAI is closely linked to other evidence regarding the nature of the instructional programs. Data collected over the first two years of the QUASAR project provide some evidence that the desired consequences are being realized. As one might expect, the first year of the QUASAR project saw an uneven and incomplete implementation on instructional innovations and improvements. Data collected in the project through classroom observations and interviews with teachers, principals and resource partners, showed that teachers were struggling to develop the necessary knowledge and pedagogical skills in order to change their instructional practices. These data suggested that

less was actually changing in classrooms than had been planned. The QCAI data collected during that year were compatible with the picture painted by other project data, in that there was little or no improvement in student performance between the Fall and Spring administrations that year. In the second year, however, the picture painted by the observational and interview data suggested much greater implementation of the intended instructional programs in QUASAR classrooms. The QCAI data were also consistent with this picture, since QCAI performance improved substantially. In fact, for the cohort of students tested both in Fall of the first project year and in Spring of the second project year, the percentage of students providing QCAI task responses judged to be at score level 3 or 4, which represent high quality mathematical responses, nearly doubled (from 14% to 27%).

Given the goals of QUASAR as a demonstration and research project, these findings are quite encouraging both because they indicate the fidelity of the data sources and evidence being collected in the project and because they suggest that the instructional improvement efforts are taking hold and having the desired consequences for students. The success of QUASAR depends to a great extent on the contribution its student assessment efforts make both to the continued improvement of the instructional opportunities provided to students and the production of evidence of positive impact. The experience of QUASAR in balancing the perspectives of mathematical content quality, psychometric technical quality, and equity and fairness in its assessment efforts should prove valuable to others, since the QCAI is a prototypic example of a new type of mathematics assessment that can be used not only to evaluate program quality but also to improve it.

References

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education (1985). Standards for educational and psychological testing. Washington, DC: Authors.

Birman, B. F. (1987). The current operation of the Chapter 1 program: Final report for the national assessment of Chapter 1. Washington, DC: Office of Educational Research and Improvement, U. S. Department of Education.

Bourque, M. L., & Garrison, H. H. (1991). The levels of mathematics achievement: Initial performance standards for the 1990 NAEP mathematics assessment, Vol. 1: National and state summaries. Washington, DC: National Assessment Governing Board.

California State Department of Education (1989). A question of thinking: A first look at students' performance on open-ended questions in mathematics. Sacramento, CA: Author.

Cannell, J. J. (1988). National norm-referenced elementary achievement testing in America's public schools: How all fifty states are above the national average. Educational Measurement: Issues and Practice, 7, 5-9.

Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 13-104). New York: American Council on Education.

Congressional Budget Office. (1987, August). Educational achievement: Explanations and implications of recent trends. Washington, DC: U. S. Government Printing Office.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Dossey, J. A., Mullis, I. V. S., Lindquist, M. M., & Chambers, D. L. (1988). The mathematics report card. Princeton, NJ: National Assessment of Educational Progress.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4(4), 289-305.

Duran, R. P. (1989). Testing of linguistic minorities. In R. L. Linn (Ed.) Educational Measurement (3rd ed.) (pp. 573-587). New York: American Council on Education.

Epstein, J. L., & MacIver, D. J. (1989). Education in the middle grades: Overview of a national survey on practices and trends. Baltimore, MD: Center for Research on Elementary and Middle Schools, Johns Hopkins University.

Ernest, P. (1991). The philosophy of mathematics education. London: Falmer Press.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. Educational Researcher, 18(9), 27-42.

Garcia, E. (1991). Education of linguistically and culturally diverse students: Effective instructional practices (Educational Practice Report: 1). Santa Cruz, CA: University of California, Santa Cruz, National Center for Research on Cultural Diversity and Second Language Learning.

Grant, G. (1988). The world we created at Hamilton High. Cambridge, MA: Harvard University Press.

Kenney, P. A. & Tang, H. (1992, April). Conceptual and operational aspects of rating student responses to performance assessments. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Knapp, M. S., & Turnbull, B. J. (1990). Better schooling for the children of poverty: Alternatives to conventional wisdom (Vol I: Summary). Washington, DC: Office of Planning, Budget & Evaluation, U. S. Department of Education.

Kozol, J. (1991). Savage inequalities: Children in America's schools. New York: Crown Publishers.

Lane, S. (1993). The conceptual framework for the development of a mathematics assessment instrument for QUASAR. Educational Measurement: Issues and Practice, 12 (2).

Lane, S., & Glaser, R. (In press). Assessment in the service of learning. In De Corte, E. (Ed.) The international encyclopedia of education (2nd ed.).

Lane, S., Parke, C., & Moskal, B. (1992, April). Principles for developing performance assessments. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Lane, S., Stone, C. A., Ankenmann, R. D., & Liu, M. (In press). Reliability and validity of a mathematics performance assessment. International Journal of Educational Research.

Lieberman, A. (Ed.) (1986). Rethinking school improvement: Research, craft and concept. New York: Teachers College Press.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20,(8), 15-21.

Magone, M., Cai, J., Silver, E. A., & Wang, N. (In press). Validating the cognitive complexity and content quality of a mathematics performance assessment. International Journal of Educational Research.

Matthews, W., Carpenter, T. P., Lindquist, M. M., & Silver, E. A. (1984). The third national assessment: Minorities and mathematics. Journal for Research in Mathematics Education, 15, 165-171.

Mehrens, W. A. (1992). Using performance assessment for accountability purposes. Educational Measurement: Issues and Practice, 11(1), 3-9, 20.

Meier, D. (1992). Reinventing teaching. Teachers College Record, 93(1), 594-609.

Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments (ETS RR-92-39). Princeton, NJ: Educational Testing Service.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 13-104). New York: American Council on Education.

Mullis, I. V. S., Owen, E. H., & Phillips, G. W. (1990). Accelerating academic achievement: A summary of findings from 20 years of NAEP. Princeton, NJ: Educational Testing Service.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1991). The state of mathematics achievement: NAEP's 1990 assessment of the nation and the trial assessment of the states. Washington, DC: National Center for Education Statistics.

National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics (1991). Professional standards for teaching school mathematics. Reston, VA: Author.

National Research Council (1989). Everybody counts. Washington, DC: National Academy of Sciences.

National Research Council (1993). Measuring up: Prototypes for mathematics assessment. Washington, DC: National Academy of Sciences.

Oakes, J. (1985). Keeping track: How schools structure inequality. New Haven, CT: Yale University Press.

Oakes, J. (1990a). Multiplying inequalities: The effect of race, social class, and tracking on opportunities to learn mathematics and science. Santa Monica, CA: The Rand Corporation.

Oakes, J. (1990b). Opportunities, achievement, and choice: Women and minority students in science and mathematics. In C. B. Cazden (Ed.), Review of research in education, Volume 16 (pp. 153 - 222). Washington, DC: American Educational Research Association.

Pallas, A. M., Natriello, G., & McDill, E. L. (1989). The changing nature of the disadvantaged population: Current dimensions and future trends. Educational Researcher, 18 (5), 16-23

Pelavin, S. H., & Kane, M. (1990). Changing the odds: Factors increasing access to college. New York: The College Entrance Examination Board.

Pollack, J. M., Rock, D. A., & Jenkins, F. (1992, April). Advantages and disadvantages of constructed-response item formats in large-scale surveys. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Porter, A. C. (1989). A curriculum out of balance: The case of elementary school mathematics. Educational Researcher, 18(5), 9-15.

Resnick, L. B. (1987). Education and learning to think. Washington, DC: National Academy of Sciences.

Resnick, L. B. (1988). Treating mathematics as an ill-structured discipline. In R. I. Charles & E. A. Silver (Eds.), Research agenda for mathematics education: Vol. 3. The teaching and assessing of mathematical problem solving (pp. 31-60). Hillsdale, NJ/Reston, VA: Lawrence Erlbaum Associates/National Council of Teachers of Mathematics.

Robitaille, D., & Garden, R. (Eds.) (1989). The IEA study of mathematics II: Contexts and outcomes of school mathematics. Oxford, England: Pegamon Press.

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In J. F. Voss, D. N. Perkins & J. W. Segal (Eds.), Informal reasoning and education (pp. 311-343). Hillsdale, NJ: Lawrence Erlbaum Associates

Secada, W. G. (1992). Race, ethnicity, social class, language, and achievement in mathematics. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 623-660). New York: Macmillan.

Silver, E. A. (1992a). Assessment and mathematics education reform in the United States. International Journal of Educational Research, 17(5), 489-502.

Silver, E. A. (1992b, August). Mathematical thinking and reasoning for all students: Moving from rhetoric to reality. Paper presented at the Seventh International Congress on Mathematical Education, Qu

ebec, Canada.

Silver, E. A., Kilpatrick, J., & Schlesinger, B. (1990). Thinking through mathematics. New York: College Entrance Examination Board.

Silver, E. A., & Lane, S. (1993). Assessment in the context of mathematics instruction reform: The design of assessment in the QUASAR project. In M. Niss (Ed.), Assessment in mathematics education and its effects (pp. 59-70). London: Kluwer Academic.

Silver, E. A., Shapiro, L. J., & Deutsch, A. (1993). Sense-making and the solution of division problems involving remainders: An examination of students' solution processes and their interpretations of solutions. Journal for Research in Mathematics Education, 24(2), 117-135.

Silver, E. A., Smith, M. S., & Nelson, B. S. (In press). The QUASAR project: Equity concerns meet mathematics education reform in the middle school. In E. Fennema, W. Secada, & L. Byrd (Eds.), New directions for equity in mathematics education.

Stein, M. K., Grover, B. W., & Silver, E. A. (1991a, October). A methodology for studying teacher change within a mathematics education reform project. Paper presented at a presession, "Methodologies for studying teacher change in the reform of school mathematics," prior to the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Blacksburg, VA.

Stein, M. K., Grover, B. W., & Silver, E. A. (1991b). Changing instructional practice: A conceptual framework for capturing the details. In R. G. Underhill (Ed.), Proceedings of the Thirteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 36-41). Blacksburg, VA: Virginia Polytechnic & State University.

Stodolsky, S. (1988). The subject matters: Classroom activity in mathematics and social studies. Chicago: University of Chicago Press.

Table 1

Specifications for the QUASAR Cognitive Assessment Instrument

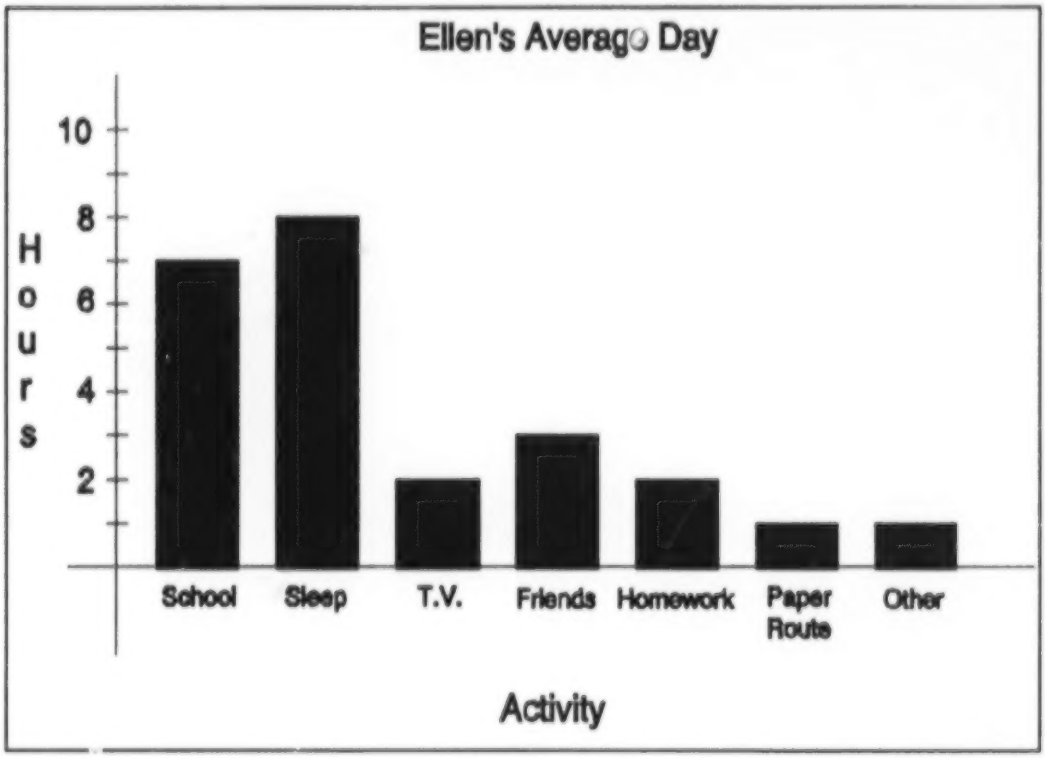
Component	Subcomponents
Content	Number and Operation (whole numbers, decimals, fractions, integers, ratios, proportions, percentages)
	Estimation (computational and measurement)
	Patterns (numerical and geometric)
	Prealgebra Skills
	Geometry
	Measurement
	Probability
	Statistics
Cognitive Processes	Understanding and representing mathematical problems
	Discerning mathematical relationships
	Organizing information
	Using strategies, procedures, and heuristic processes
	Formulating conjectures
	Evaluating the reasonableness of answers
	Generalizing results
	Justifying answers or procedures
	Communicating mathematical ideas
Modes of Representations	Text
	Pictorial display
	Graph
	Table
	Arithmetic and algebraic expressions
Real World Context	Embedded
	Not embedded

Figure 1

Sample Assessment Tasks

Task 1 - Mathematical Content: Data Analysis/Statistics (graph reading and interpretation)

Below is a graph of the activities which make up an average school day for Ellen.



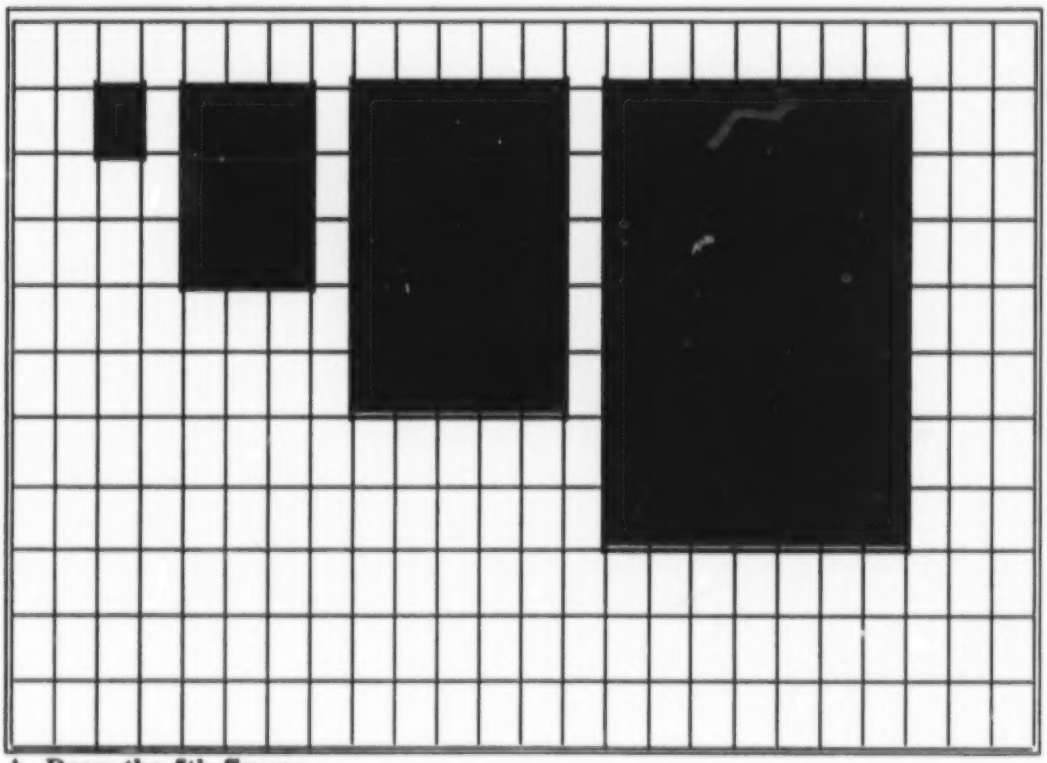
Write a story about one day in Ellen's life based on the information in the graph.

Desired Responses: We would expect a student to write a story that incorporates both dimensions of the graph -the various activities in Ellen's day and how long each one takes. The student should express the situation in realistic terms such as placing the events in chronological order. For example:

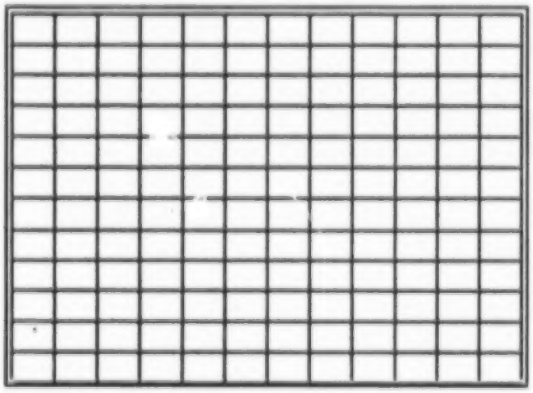
> Ellen woke up one hour early to do her paper route before going to school. After school, she and her friends went to the mall for three hours. Ellen wanted to watch her favorite shows on T.V. that night, so when she got home from the mall, she went right to her room to do her homework. It took her two hours to do her homework which left her plenty of time to see her shows. After watching T.V. for two hours she got ready for bed, tired after a long day.

Task 2 - Mathematical Concept: Pattern recognition

Look at the following pattern of figures:



A. Draw the 5th figure:



B. Describe the pattern.

Desired Responses: [illegible] described by the student would be acceptable and describe the pattern.

- It is a pattern of squares with odd sides - 1, 3, 5, 7, 9, 11.... OR
- In the pattern, you add 2 rows and 2 columns to each square to get the next square.

BEST COPY AVAILABLE

Task 3 - Mathematical Content: Numbers and Operations

The table below shows the cost for different bus fares.

BUSY BUS COMPANY
FARES

One Way	$1.00
Weekly Pass	$9.00

Yvonne is trying to decide whether she should buy a weekly bus pass. On Monday, Wednesday and Friday she rides the bus to and from work. On Tuesday and Thursday she rides the bus to work, but gets a ride home with her friends.

Should Yvonne buy a weekly bus pass?
Explain your answer.

Desired Responses: We would expect that the student's response would show evidence of a clear reasoning process. We would expect a student to answer "no" and to provide an explanation. For example:

> Yvonne takes the bus eight times in the week, and this would cost $8.00. The bus pass costs $9.00, so Yvonne should not buy the bus pass.

We would take into account, however, other plausible answers. The student may answer "yes" and provide a logical reason. For example:

> Yvonne should buy the bus pass because she rides the bus eight times for work and this costs $8.00. If she rides the bus on weekends (to go shopping, etc.), it would cost $2.00 or more, and that would be more than $9.00 altogether. So she will save money with the bus pass.

Figure 2

Block Task and Sample Student Responses at Each Score Level

4 level --This student showed complete and correct work, and provided the correct answer.

3 level -- This student showed complete and clear work with a minor error - a set of blocks was missed when the blocks were grouped in groups of 3.

2 level -- In this student's work, the total number of blocks was the same when the blocks were grouped in groups of 2 and 4, and there was a remainder of 1 in both of the groupings. However, the work does not satisfy the other constraint of the problem - the blocks should also be grouped in groups of 3 with one left over. Consequently, the student's answer is incorrect.

1 level -- This student formed one group of 13 blocks. There was no indication that the blocks were partitioned in groups of 2, 3, or 4.

0 level -- This student's work showed no understanding of the constraints of the problem.

Appendix A
General Rubric Components

I. Mathematical knowledge
   - A. Understanding of relevant mathematical concepts and principles
   - B. Use of mathematical terminology and notation
   - C. Procedural knowledge (i.e., execution of arithmetic algorithms)

II. Strategic knowledge
   - A. Construction of an appropriate problem representation
     1. Use of outside information of a formal and/or informal nature
     2. Understanding of the relationships among problem elements
     3. Relation of answer to what the question is asking
   - B. Appropriateness of solution strategy
   - C. Systematicity of solution strategy
     1. Systematic premise-driven approach
     2. Random, hit-or-miss approach

III. Communication
   - A. For problems involving mathematical arguments/justification
     1. Integrates mathematical ideas and makes conjectures, develops convincing arguments
        - a. appropriate mode (e.g.,written, pictorial, graphical, or symbolic)
        - b. structure
          - i. logically complete
          - ii. makes appropriate use of examples and counter-examples
     2. Specific justification types
        - a. justification of produced answer
        - b. justification of given proposition
   - B. For problems involving a procedural description

     1.Models situations using written, pictorial, graphical, and symbolic methods
        - a. appropriate mode and structure

     2. Specific solution types
        - a. description of solution process

     description of pattern
   - C. For all problems involving written explanations and descriptions, grammatical and spelling errors are ignored.

Appendix B
General Rubric

Score Level 4

Mathematical knowledge: Shows understanding of the problem's mathematical concepts and principles; uses appropriate mathematical terminology and notations; and executes algorithms completely and correctly.

Strategic knowledge: May use relevant outside information of a formal or informal nature; identifies all the important elements of the problem and shows understanding of the relationships between them; reflects an appropriate and systematic strategy for solving the problem; and gives clear evidence of a solution process, and solution process is complete and systematic.

Communication: Gives a complete response with a clear, unambiguous explanation and/or description; may include an appropriate and complete diagram; communicates effectively to the identified audience; presents strong supporting arguments which are logically sound and complete; may include examples and counter-examples.

Score Level 3

Mathematical knowledge: Shows nearly complete understanding of the problem's mathematical concepts and principles; uses nearly correct mathematical terminology and notations; executes algorithms completely; and computations are generally correct but may contain minor errors.

Strategic knowledge: May use relevant outside information of a formal or informal nature; identifies the most important elements of the problems and shows general understanding of the relationships between them; and gives clear evidence of a solution process, and solution process is complete or nearly complete, and systematic.

Communication: Gives a fairly complete response with reasonably clear explanations or descriptions; may include a nearly complete, appropriate diagram; generally communicates effectively to the identified audience; presents supporting arguments which are logically sound but may contain some minor gaps.

Score Level 2

Mathematical knowledge: Shows understanding of some of the problem's

mathematical concepts, and principles; and may contain serious computational errors.
Strategic knowledge: Identifies some important elements of the problems but shows only limited understanding of the relationships between them; and gives some evidence of a solution process, but solution process may be incomplete or somewhat unsystematic.
Communication: Makes significant progress towards completion of the problem, but the explanation or description may be somewhat ambiguous or unclear; may include a diagram which is flawed or unclear; communication may be somewhat vague or difficult to interpret; and arguments may be incomplete or may be based on a logically unsound premise.

Score Level 1

Mathematical knowledge: Shows very limited understanding of the problem's mathematical concepts, and principles; may misuse or fail to use mathematical terms; and may contain make major computational errors.
Strategic knowledge: May attempt to use irrelevant outside information; fails to identify important elements or places too much emphasis on unimportant elements; may reflect an inappropriate strategy for solving the problem; gives incomplete evidence of a solution process; solution process may be missing, difficult to identify, or completely unsystematic.
Communication: Has some satisfactory elements but may fail to complete or may omit significant parts of the problem; explanation or description may be missing or difficult to follow; may include a diagram which incorrectly represents the problem situation, or diagram may be unclear and difficult to interpret.

Score Level 0

Mathematical knowledge: Shows no understanding of the problem's mathematical concepts and principles.
Strategic knowledge: May attempt to use irrelevant outside information; fails to indicate which elements of the problem are appropriate; copies part of the problem, but without attempting a solution.
Communication: Communicates ineffectively; words do not reflect the problem; may include drawings which completely misrepresent the problem situation.